Bachelorseminar:
Ausgewählte Aspekte der Wirtschafts- und Sozialstatistik

Explorative Faktorenanalyse

Simon Reitzner

Betreuerin: Eva Endres, M.Sc.

LMU München, Institut für Statistik

19.06.2015

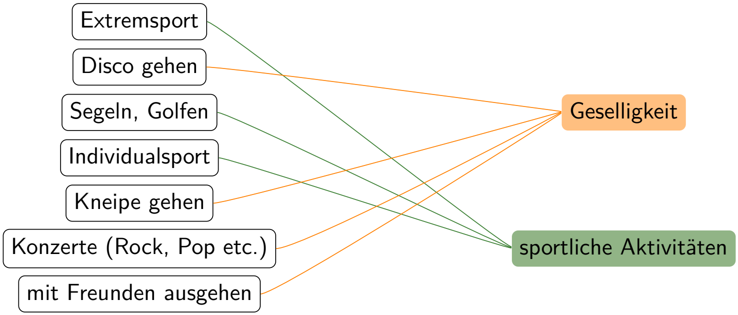
Ziele der explorativen Faktorenanalyse

- Entdecken von übergeordneten latenten Variablen
- Dimensionsreduktion

Beispiele für die Anwendung der EFA

- Big Five/Fünf-Faktorenmodell
  Beschreibung der Persönlichkeit durch 5 Faktoren: Neurotizismus, Extraversion, Offenheit für Erfahrungen, Gewissenhaftigkeit und Verträglichkeit
- Intelligenztests: globale Intelligenz, sprachliches Intelligenz, mathematische Intelligenz

Zuordnung verschiedener Freizeitaktivitäten zu Faktoren:



Grundidee

$$z_{ij} = l_{j1} f_{i1} + ... + l_{jl} f_{il} + ... + l_{pk} f_{ik} + e_{ij}$$

z_{ij}	standardisierter Wert einer Beobachtung/Person i auf eine Variable j
l_{jl}	Ladung der Variable j auf Faktor l
f_{il}	Faktorwert von Beobachtung/Person i auf Faktor l
e_{ij}	Fehlerkomponente

Matrixdarstellung

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^T + \mathbf{E}$$

Z	standardisierte Datenmatrix
L	Ladungsmatrix
F	Matrix der Faktorenwerte
E	spezifische Faktoren

Modell

$$\mathbb{E}(\mathbf{f}) = 0, \quad \mathbb{E}(\mathbf{e}) = 0$$

Faktoren korrelieren nicht mit den spezifischen Faktoren:

$$Cov(\mathbf{f}, \mathbf{e}) = 0$$

unkorrelierte spezifische Faktoren:

$$Cov(\mathbf{e}) = \text{diag}(v_1^2, ..., v_p^2) = \mathbf{V}$$

für die Ladungsmatrix gilt:

$$Cov(\mathbf{z}, \mathbf{f}) = \mathbf{L}$$

orthogonale und oblique Faktoren

orthogonales Modell:

$$Cov(\mathbf{f}) = \mathbf{I}$$

$\longrightarrow$ unabhängige Faktoren

obliques Modell:

$$Cov(\mathbf{f}) = \boldsymbol{\Phi}$$

$\longrightarrow$ abhängige Faktoren

Schätzproblem

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^T + \mathbf{E}$$

bekannt:	Datenmatrix $\mathbf{Z}$ mit p Variablen
unbekannt:	$\mathbf{F}$, $\mathbf{L}$, $\mathbf{E}$ und die Faktorenanzahl k

Fundamentaltheorem der Faktorenanalyse

Folgende Varianzzerlegung wird angestrebt:

$$\mathbf{\Sigma} = Cov(\mathbf{z}) = \mathbf{L}\mathbf{L}^T + \mathbf{V}$$

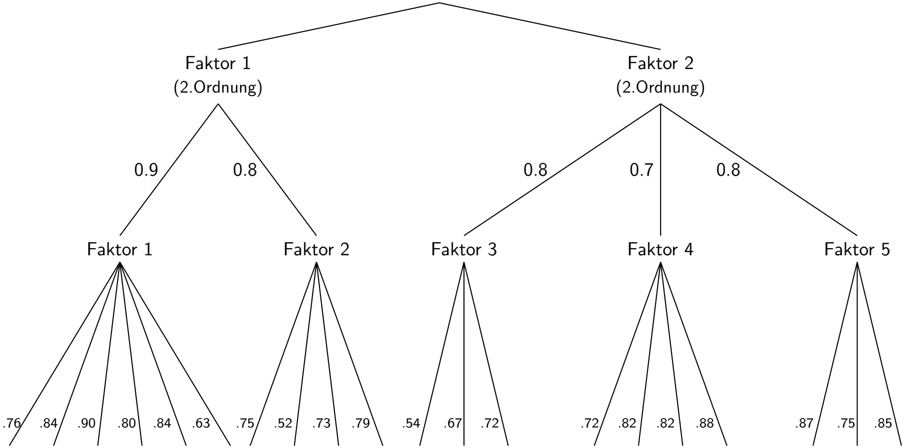
$\mathbf{\Sigma}$ wird durch die empirische Korrelationsmatrix $\mathbf{R}$ geschätzt:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T + \mathbf{V}$$

Schritte der explorativen Faktorenanalyse

- Berechnung der Korrelationsmatrix
- Faktorenextraktion
  - Bestimmen der Faktorenanzahl
  - Schätzen der Kommunalitäten
  - Berechnen der Ladungsmatrix
- Faktorenrotation
- Berechnung der Faktorenwerte

Struktur des Datensatzes



1. **Faktorenextraktion**

2. Rotation der Faktoren

3. Berechnung der Faktorenwerte

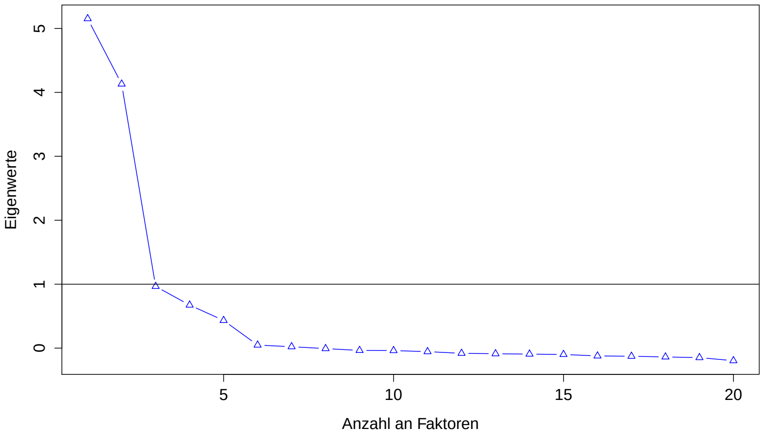
Bestimmen der Faktorenanzahl

- Eigenwertkriterium
- Scree-Test
- Parallelanalyse

Eigenwertkriterium

- Varianz, die die jeweiligen Faktoren erklären, entspricht den Eigenwerten von $\mathbf{R}$
- Faktoren bedeutsam, sie mehr als die Varianz einer standardisierten Variable erklären
- Anzahl der Faktoren entspricht Anzahl an Eigenwerten größer 1

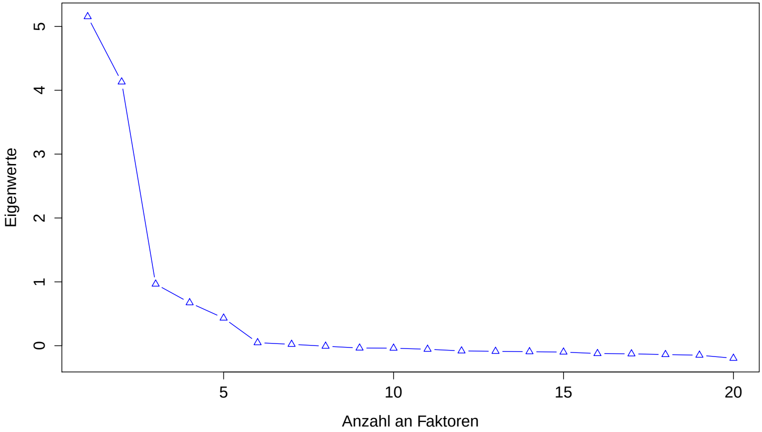
Eigenwertkriterium



Scree-Test

- grafische Bestimmung der Anzahl
- Einzeichnen der Eigenwerte in einen Plot
- Anzahl an Faktoren entspricht der Anzahl an Eigenwerten vor dem „Knick“

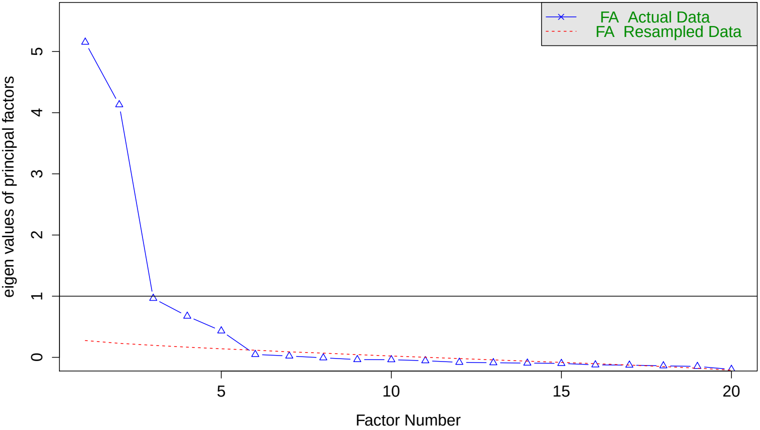
Scree-Plot



Parallelanalyse

- Vergleich der Eigenwerte mit Bootstrap-Eigenwerten
- Eigenwerte über den simulierten, zufälligen Eigenwerten sind auffällig
  $\longrightarrow$ extrahieren!

Parallelanalyse



Spektralzerlegung

Sei $\mathbf{A}$ eine symmetrische $(p \times p)$-Matrix mit $rg(\mathbf{A}) = r$. Dann existiert eine $(p \times r)$-Matrix $\mathbf{P}$, sodass gilt:

$$\mathbf{A} = \mathbf{P} \operatorname{diag}(\lambda_1, \ldots, \lambda_r) \mathbf{P}^T$$

angewendet bei der Faktorenanalyse:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T = \mathbf{P} \operatorname{diag}(\lambda_1, \ldots, \lambda_r) \mathbf{P}^T$$

für die Ladungsmatrix $\mathbf{L}$ gilt:

$$\mathbf{L} = \mathbf{P} \operatorname{diag}(\sqrt{\lambda}_1, \ldots, \sqrt{\lambda}_r)$$

Schätzmethoden des Faktorenmodells

- Hauptkomponentenmethode
- Hauptfaktorenanalyse

Hauptkomponentenmethode

Modellgleichung:

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^T$$

Zerlegung der Korrelationsmatrix:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T$$

- keine spezifischen Faktoren
- entspricht nicht der ursprünglichen Idee einer Faktorenanalyse

Hauptfaktorenanalyse

Modellgleichung:

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^T + \mathbf{E}$$

Zerlegung der Korrelationsmatrix:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T + \mathbf{V} \quad \Longleftrightarrow \quad \mathbf{R} - \mathbf{V} = \mathbf{L}\mathbf{L}^T \quad \Longleftrightarrow \quad \mathbf{R_h} = \mathbf{L}\mathbf{L}^T$$

$\mathbf{R_h}$: reduzierte Korrelationsmatrix

Unterschied zwischen Hauptfaktorenanalyse und Hauptkomponentenmethode

Hauptkomponentenmethode:

$$\mathbf{R} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$$

Hauptfaktorenanalyse:

$$\mathbf{R_h} = \begin{pmatrix} h_1^2 & & \\ & \ddots & \\ & & h_p^2 \end{pmatrix}$$

Kommunalitäten

- Varianzanteil einer Variable, der durch die Faktoren erklärt wird
- iterative Schätzung
- Startkommunalitäten:
  - quadrierter multiple Korrelationskoeffizient
  - betragsmäßig größte Korrelationskoeffizient zweier Variablen

Iterative Schätzung der Kommunalitäten

- Einsetzen der Startkommunalitäten in $\mathbf{R}$
- Berechnen der Ladungsmatrix $\mathbf{L}$ mit k Eigenvektoren und Eigenwerten
- Berechnung von $\mathbf{L}\mathbf{L}^T = \mathbf{R_h}$
- Diagonale von $\mathbf{R_h}$ enthält Kommunalitäten für den 2.Iterationsschritt

1. Faktorenextraktion

2. Rotation der Faktoren

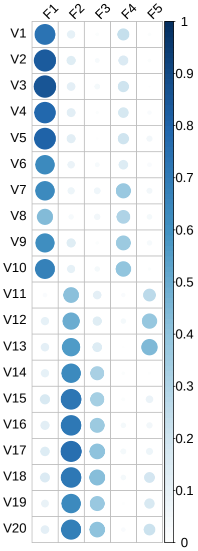
3. Berechnung der Faktorenwerte

Rotation der Faktoren

Ziel der Rotation:

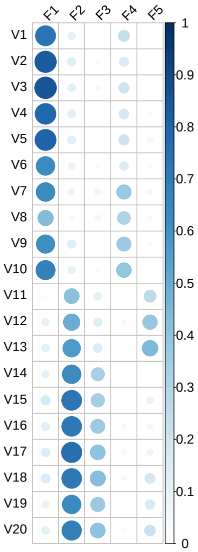
- eindeutige Zuordnung
- einfachere Interpretation

unrotierte Ladungsmatrix

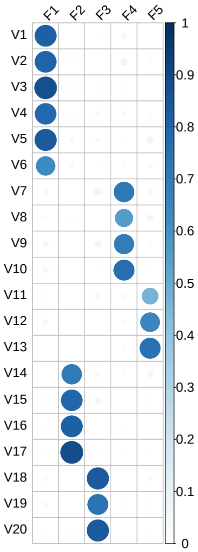


Ladungsmatrizen

unrotiert:

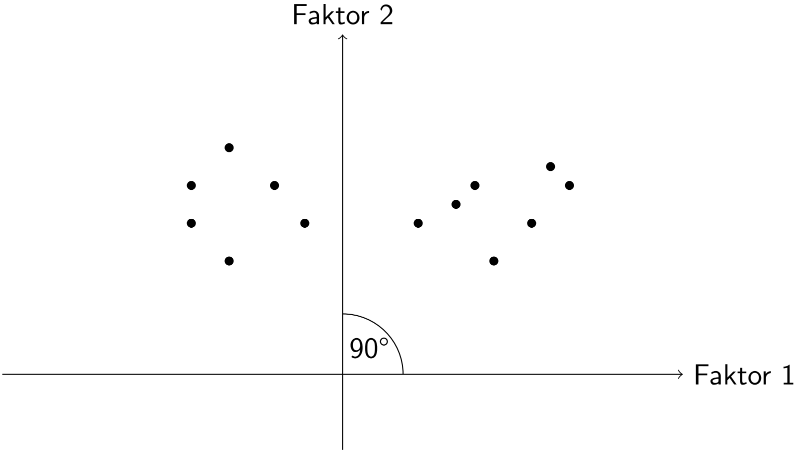


rotiert:



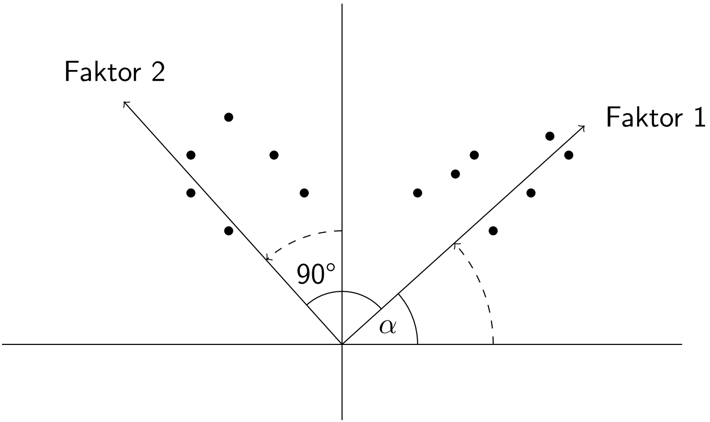
Rotation

Faktoren vor der Rotation:



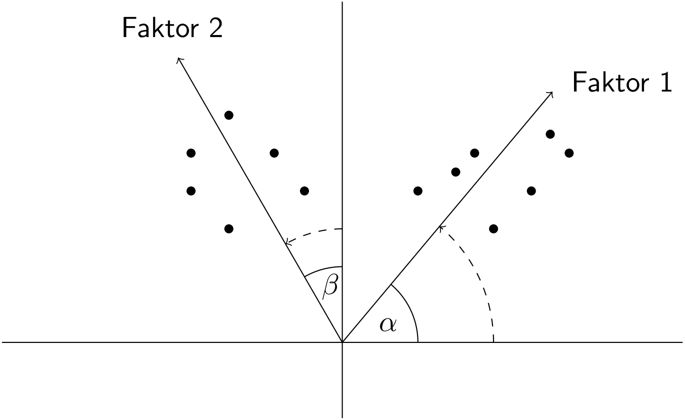
Rotation

Faktoren nach orthogonaler Rotation:



Rotation

Faktoren nach obliquer Rotation:



Rotation

Gesucht ist eine Rotationsmatrix $\mathbf{M}^\alpha$ mit Rotationswinkel α:

$$\mathbf{L} \cdot \mathbf{M}^\alpha = \tilde{\mathbf{L}}$$

Für 2 orthogonale Faktoren:

$$\mathbf{L} \cdot \mathbf{M}^\alpha = \begin{pmatrix} l_{11} & l_{12} \\ \vdots & \vdots \\ l_{p1} & l_{p2} \end{pmatrix} \cdot \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} \tilde{l}_{11} & \tilde{l}_{12} \\ \vdots & \vdots \\ \tilde{l}_{p1} & \tilde{l}_{p2} \end{pmatrix} = \tilde{\mathbf{L}}$$

Rotationsmethoden

- orthogonale Rotationen
  - Varimax:
  - Quartimax:
- oblique Rotationen
  - Promax:
  - Oblimax:

1. Faktorenextraktion

2. Rotation der Faktoren

3. Berechnung der Faktorenwerte

Faktorenwerte

- Ausprägungen der Personen auf die Faktoren
- Schätzverfahren:
  - Summenscores
  - Bartlett-Methode
  - Anderson-Rubin-Methode
  - Ten Berge

Vielen Dank für Ihre Aufmerksamkeit!

Literaturverzeichnis I

[1] J. Bortz and C. Schuster. *Statistik für Human- und Sozialwissenschaftler. Lehrbuch mit Online-Materialien*. Springer, S.415, 2010.

[2] F. Brosius. *SPSS 19*. mitp, S.787, 2011.

[3] A. Bühl. *SPSS 16: Einführung in die moderne Datenanalyse*. Pearson Studium, S.527, 2008.

[4] M. Bühner. *Einführung in die Test- und Fragebogenkonstruktion*. Pearson Studium, S.296-299, 2011.

[5] C. DiStefano, M. Zhu, and D. Mindrila. Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20): S.1–11, 2009.

Literaturverzeichnis II

[6] B. Everitt and T. Hothorn. *An Introduction to Applied Multivariate Analysis with R*. Springer New York, S.141, 2011.

[7] L. Fahrmeir, W. Brachinger, A. Hamerle, and G. Tutz. *Multivariate statistische Verfahren*. de Gruyter, S.640-693, 1996.

[8] T. Fehr. Big Five: Die fünf grundlegenden Dimensionen der Persönlichkeit und ihre dreißig Facetten. `www.i-p-p-m.de/Das_Big-Five_Modell.pdf`. aufgerufen am: 17.06.2015.

[9] J. W. Grice. Computing and evaluating factor scores. *Psychological Methods*, 6(4): S.430–450, 2001.

[10] J. L. Horn. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2): S.179–185, 1965.

Literaturverzeichnis III

[11] H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3): S.287–200, 1958.

[12] L. Knüsel. Factor analysis: Chisquare as rotation criterion. `http://epub.ub.uni-muenchen.de/6350/1/tr040.pdf`, 2008. aufgerufen am: 25.05.2015.

[13] M. Noack. Faktorenanalyse. `www.uni-due.de/imperia/md/content/soziologie/stein/faktorenanalyse.pdf`, 2007. aufgerufen am: 20.05.2015.

[14] W. Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2015. R package version 1.5.4.

Literaturverzeichnis IV

[15] O. Walter. Intelligenz. www.verhaltenswissenschaft.de/Psychologie/Personlichkeit/Intelligenz/intelligenz.htm#globalmodell. aufgerufen am: 17.06.2015.